International Journal of Pharmaceutics 457 (2013) 310–322

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

journal homepage: www.elsevier.com/locate/ijpharm

The STEP (Safety and Toxicity of Excipients for Paediatrics) database: Part 2 – The pilot version

Smita Salunke[a,*,1], Barbara Brandys[b,1], George Giacoia[c,1], Catherine Tuleu[a,1]

[a] Department of Pharmaceutics and Centre for Paediatric Pharmacy Research, UCL School of Pharmacy, London, United Kingdom
[b] National Institutes of Health (NIH) Library, Division of Library Services, Office of Research Services, National Institutes of Health, Bethesda, MD, USA
[c] Obstetric and Pediatric Pharmacology Branch (OPPB), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), USA

ARTICLE INFO

Article history:
Received 19 August 2013
Accepted 16 September 2013
Available online 23 September 2013

Keywords:
Excipients
Children
Database
Toxicity
Adverse effects
Chemicals

ABSTRACT

The screening and careful selection of excipients is a critical step in paediatric formulation development as certain excipients acceptable in adult formulations, may not be appropriate for paediatric use. While there is extensive toxicity data that could help in better understanding and highlighting the gaps in toxicity studies, the data are often scattered around the information sources and saddled with incompatible data types and formats.

This paper is the second in a series that presents the update on the **S**afety and **T**oxicity of **E**xcipients for **P**aediatrics ("STEP") database being developed by Eu-US PFIs, and describes the *architecture* data fields and functions of the database. The STEP database is a user designed resource that compiles the safety and toxicity data of excipients that is scattered over various sources and presents it in one freely accessible source. Currently, in the pilot database data from over 2000 references/10 excipients presenting preclinical, clinical, regulatory information and toxicological reviews, with references and source links. The STEP database allows searching "FOR" excipients and "BY" excipients. This dual nature of the STEP database, in which toxicity and safety information can be searched in both directions, makes it unique from existing sources. If the pilot is successful, the aim is to increase the number of excipients in the existing database so that a database large enough to be of practical research use will be available. It is anticipated that this source will prove to be a useful platform for data management and data exchange of excipient safety information.

1. Background

The development of an appropriate paediatric formulation is a complex task that requires multiple considerations. Amidst the several challenges, such as acceptable palatability, age appropriateness, dosage forms and dosing devices, is the selection of "suitably safe excipients" (Maldonado and Schaufelberger, 2011). The screening and careful selection of excipients is a critical step in paediatric formulation development as certain excipients acceptable in adult formulations, may not be appropriate for paediatric use, e.g. ethanol in oral liquids or benzyl alcohol in intravenous formulations for neonates. During infancy and childhood, there is rapid growth and development with age-related changes in the various organs, body composition, protein binding, active transport mechanisms and metabolic pathways. Therefore, paediatric patients and in particular neonates and infants, may not be able to metabolize or eliminate an ingredient in the same manner as adult and so in some instances this may lead to deleterious side effects (CHMP, 2006).

The toxicity of excipients may differ between adult and paediatric patients and across the paediatric sub-sets e.g. preterm

Abbreviations: ACToR, Aggregated Computational Toxicology Resource; ADI, Acceptable Daily Intake; AFSSAPS, Agence française de sécurité sanitaire des produits de santé (French Agency for the Safety of Health Products); CAS, Chemical Abstracts Service; CCOHS, Canadian Centre for Occupational Health and Safety; EFSA, European Food Safety Authority; ESIS, European chemical Substances Information System; ESNEE, European Study of Neonatal Excipient Exposure; EuPFI, European Paediatric Formulation Initiative; FDA, Food and Drug Administration; GRAS, Generally Regarded As Safe; IPCS, International Programme on Chemical Safety; IT, Information Technology; IUPAC, International Union of Pure And Applied Chemistry; JECFA, Joint FAO/WHO Expert Committee on Food Additives; NLM, National Library of Medicine; NRC, National Research Council; NTP, National Toxicology Program database; PK/PD, Pharmacokinetics/Pharmacodynamics; SmPC, Summary of Product Characteristics; STEP, Safety and Toxicity of Excipients for Paediatrics; URL, Uniform Resource Locator; US PFI, United States Paediatric Formulation Initiative; WHO, World Health Organization.

* Corresponding author at: Department of Pharmaceutics and Centre for Paediatric, Pharmacy Research (CPPR), UCL School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, United Kingdom. Tel.: +44 2077535846; fax: +44 2077535942.
E-mail addresses: S.Salunke@ucl.ac.uk, smitasalunkey@yahoo.co.uk (S. Salunke).

[1] On behalf of the European (Eu) & United States (US) Paediatric Formulation Initiatives (PFIs).

0378-5173
http://dx.doi.org/10.1016/j.ijpharm.2013.09.013

neonates may be at increased risk of propylene glycol-associated adverse events due to diminished ability to metabolize propylene glycol, thereby leading to accumulation and potential adverse events (FDA, 2011). Depending on the active substance and excipients, appropriate use of the medicinal product in the newborn may require a new formulation or appropriate information about dilution of an existing formulation. Although basic considerations regarding the use of a specific excipient in a medicinal product are not different for adult and paediatric medicines, the inclusion of any excipient in a paediatric medicine requires additional risk assessment focusing on potential safety concerns. Risks to children and particularly to neonates have been emphasized in several publications (Ernest et al., 2007; Fabiano et al., 2011; Ursino et al., 2011) and also recognized by the regulatory agencies (Carter, 2011).

There are many literature reports about the possible adverse effects of pharmaceutical excipients in the paediatric population, and if children are taking multiple medications they may be exposed to several potentially toxic doses of excipients while receiving routine treatment. For instance, benzyl alcohol, propylene glycol and polysorbate 80 co-administration resulted in various toxicological syndromes in paediatric populations especially neonates (Shehab et al., 2009; Kulo et al., 2012; Balistreri et al., 1986). In particular, hospitalized neonates have received medicines with potentially harmful drug formulation excipients (Lass et al., 2012). It is possible that even if the excipient is known to be potentially harmful, the daily intake will not exceed the toxic threshold due to the small quantities used in drug formulations. However, the general lack of quantitative information of the excipient amount in the Summary of Product Characteristics (SmPC) makes it difficult for the practitioner to readily making an informed decision. The extent of possible harm caused by the formulation excipients has not been established. Potentially harmful excipients need careful safety assessment and determination of the pharmacokinetic/pharmacodynamic (PK/PD) profiles in paediatrics (Allegaert et al., 2010).

An increased public awareness about the hazards and toxicity of these excipients has encouraged the development of technologies for their remediation. The way forward is to understand better and predict potential toxicities at an early stage of drug development, to gain deeper insights into the biology and underlying toxicity, and make decisions well before committing to further development and clinical trials. The combination of traditional toxicology methods with new strategies and tools for integrating high-throughput transcriptomics, proteomics, and metabolomics data could help to work towards the goal (Guengerich and MacDonald, 2007). However, the greatest challenge associated with productive evaluations is the availability of existing data. As excipients were generally considered inert and assumed to be safe in children on the basis of adult data, there are very limited studies about the extent to which children, especially premature babies and neonates, are exposed to excipients and the safety related to their use (Nunn and Williams, 2005). Also many traditional and commonly used excipients have not undergone rigorous toxicological assessment. There is still a gap between the research performed and the available evidence base in paediatrics compared to that in adults. There is an urgent need to identify and fill in the gaps in toxicological assessment of excipients. Information about the possible harm resulting from excipients and also the quantitative data regarding the excipient amounts in specific drug products should be made available to pharmacists, drug development scientists and paediatricians. This will assist the selection of appropriate medicines for the paediatric population and support the development of paediatric medicines. When excipients cannot be avoided, professionals should have access to quantitative and qualitative information that allows them to assess risk. The existing public sources do not provide complete and/or comparative information on safe use and acceptability of excipients in paediatrics. While there is extensive toxicity data that could help in better understanding and highlighting the gaps in toxicity studies, the data are often scattered around and saddled with incompatible data types, formats, databases, and analysis tools. Finally much information such as description of experimental methods, details about test, dosage regimen, etc., is often stored as free text format which is not easy to search. To be able to derive informed decisions in an efficient way, in a short time, it was necessary to integrate the safety and toxicity information of excipients from disparate resources under one umbrella and design and establish a tool for easy, quick and reliable access of this information. Thus, the development of a web-based database of **S**afety and **T**oxicity of **E**xcipients for **P**aediatrics ("STEP") has been undertaken collaboratively by European (Eu)[2] and United States (US)[3] Paediatric Formulation Initiatives (PFIs). The aim is to gather and extract in systematic way the available toxicity and safety information and organize it into a searchable system allowing data mining, visualization and analysis to discover useful information from collected data sets. This freely/publicly accessible source will help enhance the evidence base of safety and toxicity information of excipients for the pharmaceutical industry, academics, pharmacists, clinicians and regulators to make informed decisions and thus expedite paediatric drug development.

The establishment of successful data sharing strategies requires buy-in from end-users and agreement on the content, structure and format of the database. Furthermore, it is advisable to obtain a solid understanding of the original purpose of the data, because the way those data were collected and verified or validated will help shape or limit their use in the database. Hence, the pilot database development was planned after determining the needs captured through the global need assessment survey (Salunke et al., 2012). The need assessment study focussed on the factors influencing the decision that the database should be implemented and this pilot study focuses on how it should be implemented. The intention is to identify potential concerns and challenges, such as data collection strategy, data curation methods, handling data difficult to access or curate, and practical issues in data quality assessments. This information can then be used to modify the database components and reach improved final product.

Through the present pilot database the aim is to assess comprehension, acceptance, feasibility, and other factors that influence how readily the database will meet the needs captured from the end-users and how well it will facilitate their workflow

After the development phase, the pilot version of the STEP database compiling the data for 10 prioritized excipients is now available for beta testing from the EuPFI website. If the pilot is proved to be successful, the database will be expanded to a fully released database and will eventually include many excipients.

[2] The European Paediatric Formulation Initiative (EuPFI) is a consortium founded in 2007 and working in a pre-competitive way on paediatric drug formulations. Members are from academia, hospital pharmacies, pharmaceutical industry (Innovators, Generics, Contract Research Organizations (CRO), Specials and Excipient Manufacturers) with European Medicine Agency (EMA) as an observer. The main objective of the members is to resolve scientific, regulatory and technological issues associated with paediatric formulation development.

[3] The United States Paediatric Formulation Initiative (US-PFI) is a project of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The PFI was established in 2005 to address the issue of the lack of appropriate formulations in children and to use this activity as a means to improve paediatric formulations.

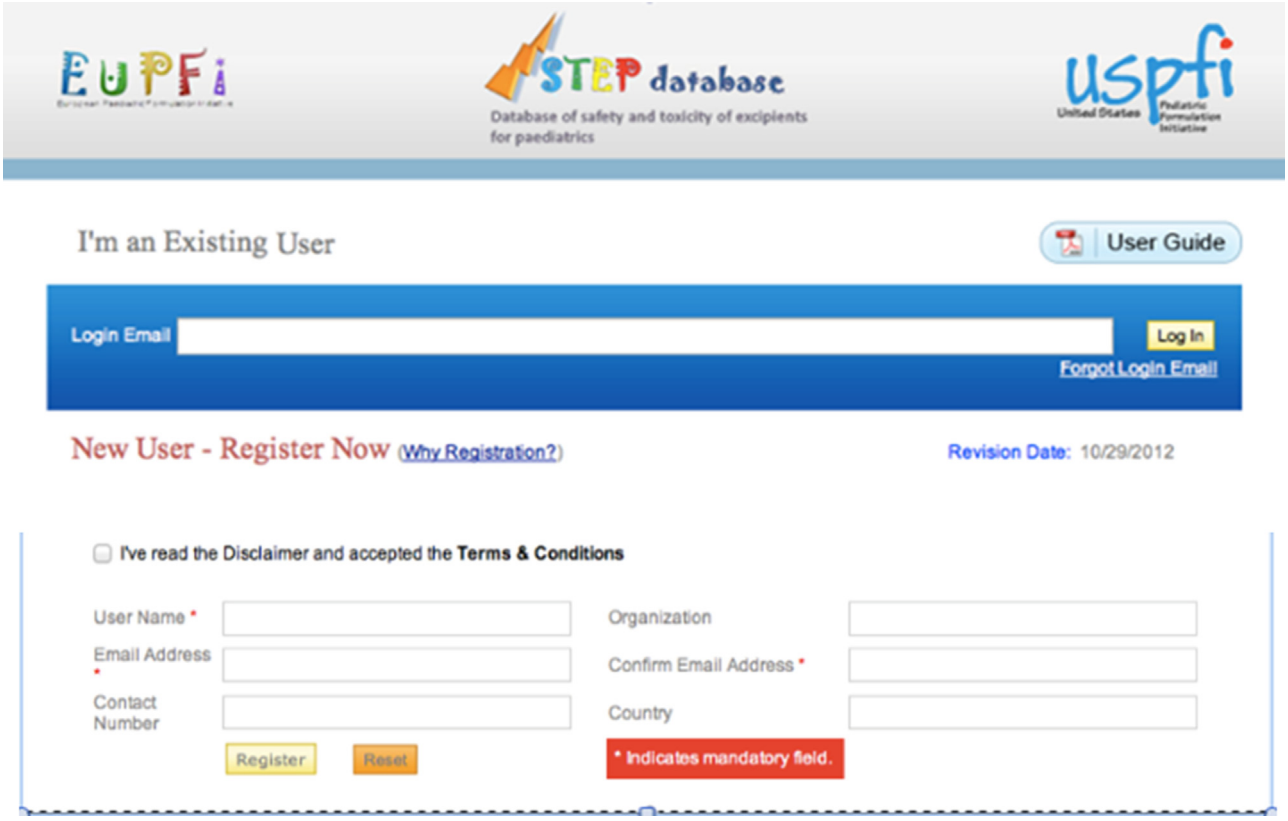


Fig. 1. Screen shot of the registration page.

2. Technical details

2.1. Functional requirements

The STEP Database is a "user-designed" resource for "users" and compiles the safety and toxicity data of excipients that is scattered over various sources and presents it in one freely accessible source. The critical appraisal of the references is out of scope of the STEP database and so the database will not provide any interpretive comments. Hence, a user will be encouraged to access such information at the source/link provided. Also, the responsibility of interpreting and using this information appropriately will always remain with the end-users.

The following content and structure requirements were formulated from the global need assessment survey aforementioned:

1. Content:
   - Experimental details and results from in vivo animal experiments, in vitro experiments, human safety and toxicity studies and epidemiology data.
   - Regulatory information and toxicological reviews.
2. Structure:
   - Provide the information in tabular/granular format.
   - Enable viewing and filtering/refining the information to facilitate the Users' needs.
   - Create a search tool to accommodate complex queries.
   - Allow exporting outcome data into PDF, Word and Excel format as needed.
   - The design should be sufficiently flexible to allow necessary modifications and yet maintain backward compatibility for the components existing in the system.

2.2. General structure of the database

The database is organized in six sections functioning in parallel (1) General Information, (2) Clinical Data, (3) Non-Clinical Data, (4) In Vitro Data, (5) Regulatory References and (6) Reviews. Each section is detailed in paragraphs below. The STEP database will be free and publicly accessible through EuPFI website but it will require registration and authorization of the password.

3. Database construction and content

3.1. Database architecture

The database is built up on an intuitive search interface and results display. The launching page expedites sign-up process with established credentials, and provides account registration form for the first time users (Fig. 1). On successful registration the user is issued a password and subsequently can access the database with their email ID and password.

3.1.1. Search interface

The search interface offers two search modules: "Search **by** Excipient" and "Search **for** Excipients". "Search **by** Excipient" module provides navigation tools for selection of specific excipients and should be used when searching for safety and toxicity information of particular excipients. It allows for browsing by alphabetically sorted excipients' names, selecting an excipient or multiple excipients, or by browsing and selecting additional excipents' identification options:

1. Chemical name/Synonyms
2. CAS number (the unique, universal, numeric chemical substance identifiers)
3. Functionality of excipients

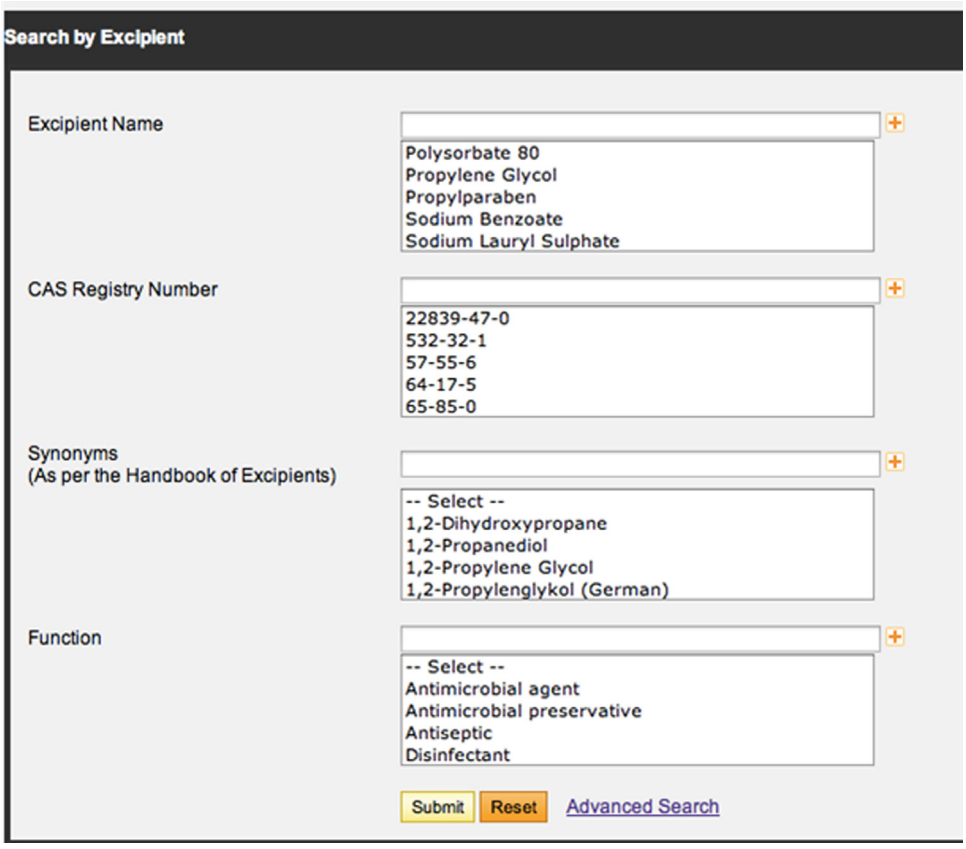


Fig. 2. Screen shot of Search by Excipients module.

Fig. 2 is the screen shot of "Search **by** Excipients" module illustrating how the user can search by the excipients name (e.g. Propylene glycol), or synonym (e.g. 1,2 propanediol), or CAS number (e.g. 57-55-6) or functionality of the excipient (e.g. Solvents, Preservatives).

The searches can be performed by either selecting the options available in the drop down menu or by typing only partial name in the "Excipient Name" search box with autocomplete/suggestion capability and selecting name from suggested list. For instance, if the user types "ben" in the excipient name box, the system will give benzoic acid, benzyl alcohol, benzyl paraben, etc. in the drop down box as indicated in Fig. 3. Fig. 4 shows how search can be performed by selecting the excipients in the drop down box. On submitting the query, the user is directed to the search results page.

The "Search **for** Excipients" module provides enhanced tools for complex queries allowing for searching for excipients associated with specific studies, effects or pharmacological functions (Fig. 5). The users can find the excipients by their adverse effect, administration age, study type and/or route of administration (e.g. which excipients are administered in infants by intravenous route).

3.1.2. Search results display

The "Search Results" from a database query are presented in tabular format (Fig. 6) and can be sorted, refined and exported according to the user needs.

Each element of the Search Result display is discussed in more details in following section

3.1.2.1. General information. This section is structured to provide the basic/descriptive information of the excipient in an easily readable format. Table format is provided to summarize the information and to help the users to screen the information at a glance (Fig. 7). The section lists identification information including Excipient's name, key Synonyms and CAS number. The identification

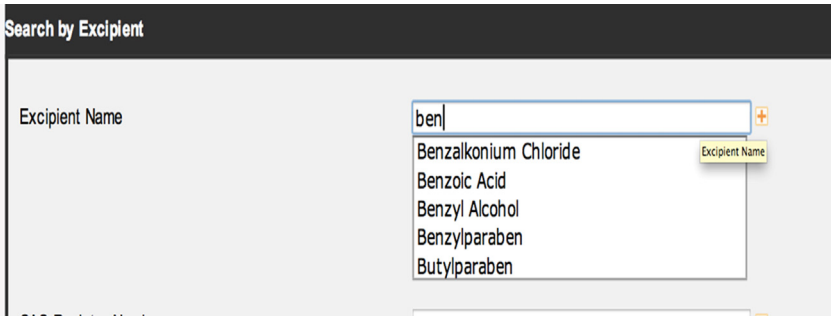


Fig. 3. Screen shot of searching the excipients by typing three letters of the excipients to be searched.

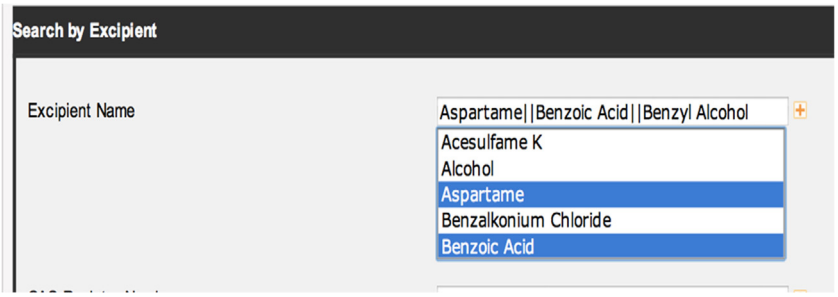


Fig. 4. Screen shot of searching the multiple excipients.

information particularly excipient chemical name, synonyms and CAS number are as according to the Handbook of Pharmaceutical Excipients (Rowe et al., 2012) for pilot database. Additional identification data and chemical and physical properties of excipients such as chemical common name, chemical formula or molecular weight might be retrieved via link to the ChemID database (Tomasulo, 2002). ChemID is a free, web-based search system that provides access to structure and nomenclature authority files used for the identification of chemical substances cited in National Library of Medicine (NLM) databases, including the TOXNET® system.

The other key information provided in the General Information section is "Regulatory status" and "Acceptable Daily Intake (ADI)". The regulatory status information is retrieved from the government websites (e.g. FDA, EMEA, AAFSAPS, WHO) and other key publicly available sources (e.g. Legacy documents). The specific data including the agency name, the description, limits for use and reference URL are extracted from the information retrieved (e.g. toxicological profiles).

The **Acceptable daily intake** or **ADI** is a measure of the amount of a specific substance (originally applied for a food additive, later also for a residue of a veterinary drug or pesticide) in food or drinking water that can be ingested (orally) without an appreciable health risk without an appreciable health risk for a life time questionable significance. ADI information is divided into two components (1) general acceptable daily intake and (2) paediatric acceptable daily intake. The database of food additives (JECFA) evaluated by the World Health Organization (WHO) suggests the ADI and has made a general exception for the use of the ADI in stating that the ADI should not be considered applicable to neonates and young infants up to 12 weeks of age (WHO, 1987). The EU Scientific Committee for Food also recommends that intake assessment of children be considered separately from that of adults because patterns of consumption are different. ADIs should cover the entire population including children. Considering this fact, the STEP database provides separate sections for general and paediatric specific ADIs. Any ADI information available for excipients from existing literature is compiled under "general acceptable daily intake" and only information gathered that is specific for paediatric population is reported in the STEP under the section for paediatric acceptable daily intake. If no paediatric ADI information is obtained from literature for any particular excipients then it is listed as "Not Available". Overall the ADI information is retrieved primarily from

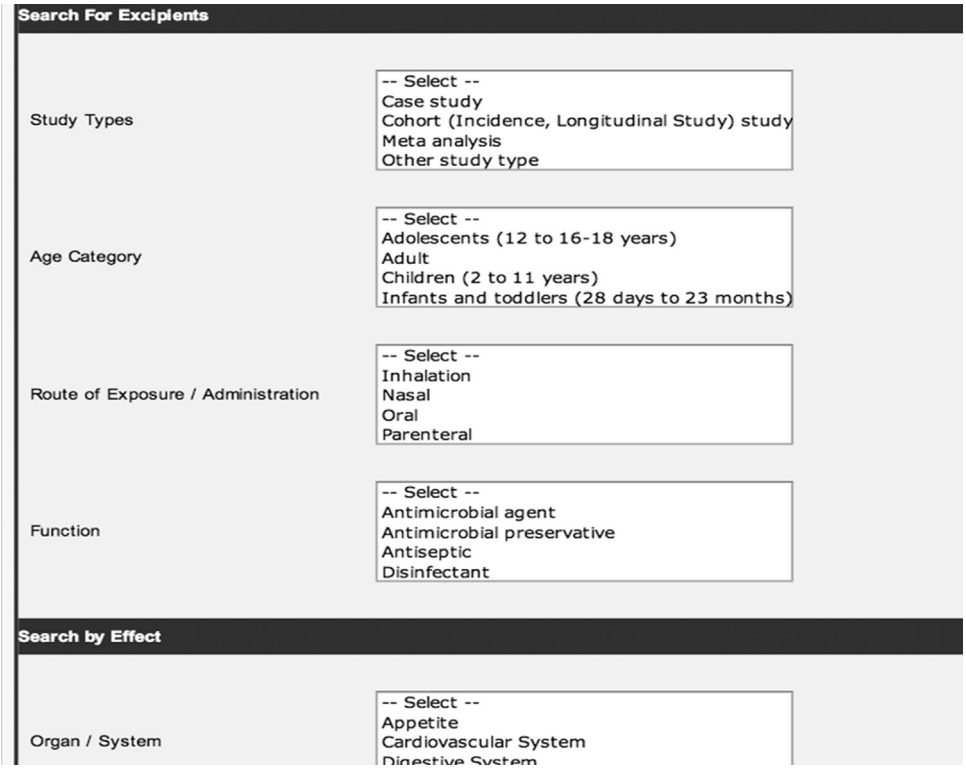
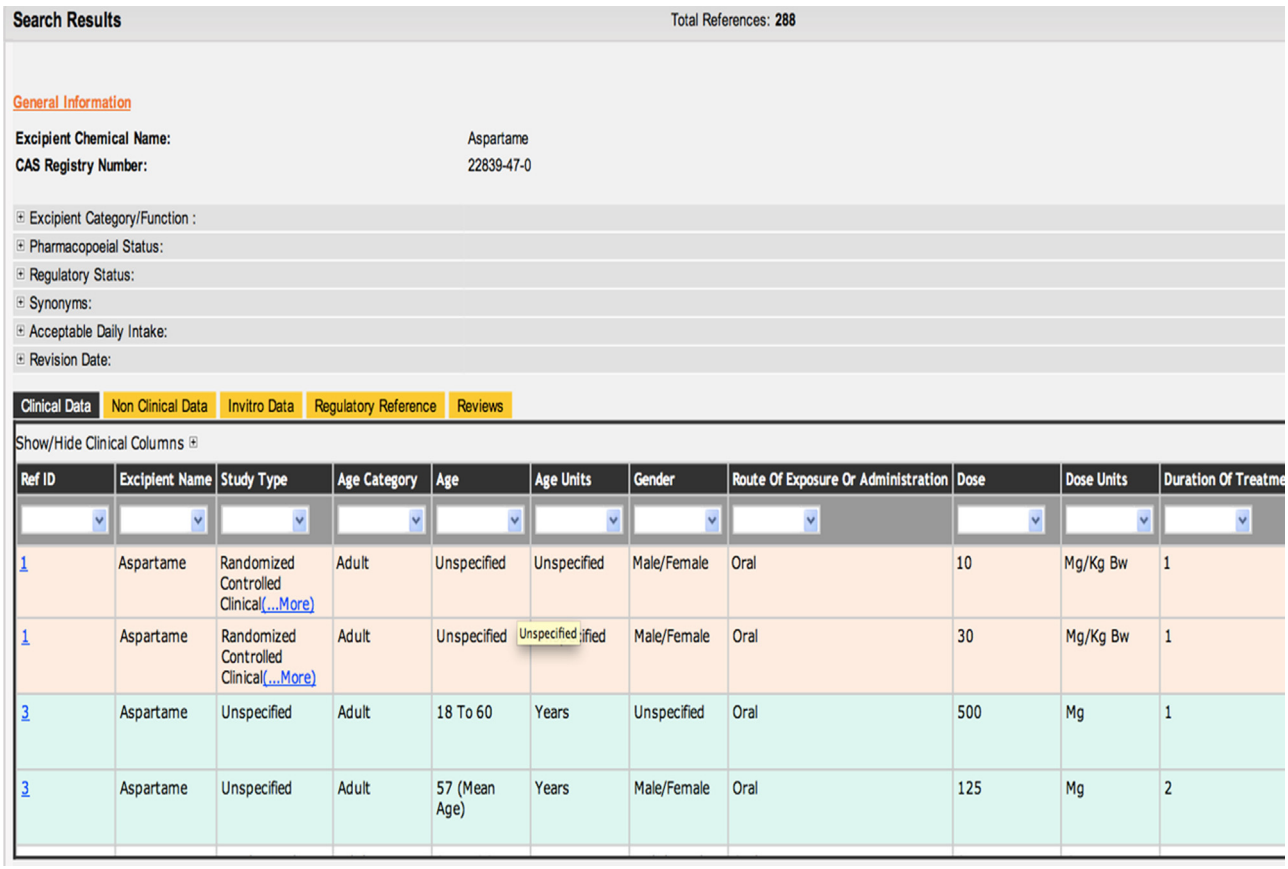


Fig. 5. Screen shot of Search for Excipients module.



Search Results Total References: **288**

General Information

Excipient Chemical Name: Aspartame
CAS Registry Number: 22839-47-0

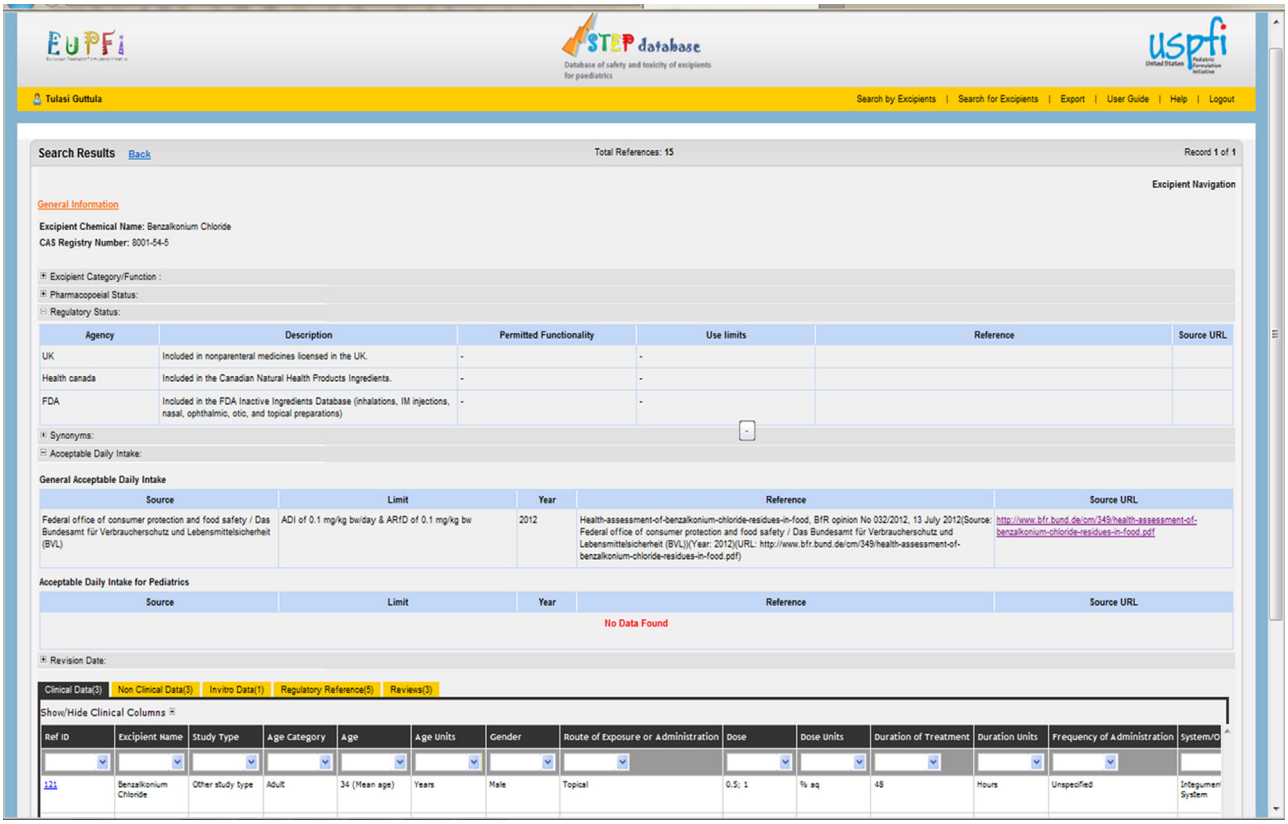
- Excipient Category/Function :
- Pharmacopoeial Status:
- Regulatory Status:
- Synonyms:
- Acceptable Daily Intake:
- Revision Date:

Clinical Data | Non Clinical Data | Invitro Data | Regulatory Reference | Reviews

Show/Hide Clinical Columns

Ref ID	Excipient Name	Study Type	Age Category	Age	Age Units	Gender	Route Of Exposure Or Administration	Dose	Dose Units	Duration Of Treatme
1	Aspartame	Randomized Controlled Clinical(...More)	Adult	Unspecified	Unspecified	Male/Female	Oral	10	Mg/Kg Bw	1
1	Aspartame	Randomized Controlled Clinical(...More)	Adult	Unspecified	Unspecified	Male/Female	Oral	30	Mg/Kg Bw	1
3	Aspartame	Unspecified	Adult	18 To 60	Years	Unspecified	Oral	500	Mg	1
3	Aspartame	Unspecified	Adult	57 (Mean Age)	Years	Male/Female	Oral	125	Mg	2

Fig. 6. Screen shot of Search Results display.



EuPFi | STEP database — Database of safety and toxicity of excipients for paediatrics | uspfi

Tulasi Guttula — Search by Excipients | Search for Excipients | Export | User Guide | Help | Logout

Search Results Back — Total References: 15 — Record 1 of 1

Excipient Navigation

General Information

Excipient Chemical Name: Benzalkonium Chloride
CAS Registry Number: 8001-54-5

- Excipient Category/Function :
- Pharmacopoeial Status:
- Regulatory Status:

Agency	Description	Permitted Functionality	Use limits	Reference	Source URL
UK	Included in nonparenteral medicines licensed in the UK.	-	-		
Health canada	Included in the Canadian Natural Health Products Ingredients.	-	-		
FDA	Included in the FDA Inactive Ingredients Database (inhalations, IM injections, nasal, ophthalmic, otic, and topical preparations)	-	-		

- Synonyms:
- Acceptable Daily Intake:

General Acceptable Daily Intake

Source	Limit	Year	Reference	Source URL
Federal office of consumer protection and food safety / Das Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL)	ADI of 0.1 mg/kg bw/day & ARfD of 0.1 mg/kg bw	2012	Health-assessment-of-benzalkonium-chloride-residues-in-food, BfR opinion No 032/2012, 13 July 2012(Source: Federal office of consumer protection and food safety / Das Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL))(Year: 2012)(URL: http://www.bfr.bund.de/cm/349/health-assessment-of-benzalkonium-chloride-residues-in-food.pdf)	http://www.bfr.bund.de/cm/349/health-assessment-of-benzalkonium-chloride-residues-in-food.pdf

Acceptable Daily Intake for Pediatrics

Source	Limit	Year	Reference	Source URL
No Data Found				

- Revision Date:

Clinical Data(3) | Non Clinical Data(3) | Invitro Data(1) | Regulatory Reference(5) | Reviews(3)

Show/Hide Clinical Columns

Ref ID	Excipient Name	Study Type	Age Category	Age	Age Units	Gender	Route of Exposure or Administration	Dose	Dose Units	Duration of Treatment	Duration Units	Frequency of Administration	System/O
121	Benzalkonium Chloride	Other study type	Adult	34 (Mean age)	Years	Male	Topical	0.5; 1	% aq	48	Hours	Unspecified	Integumen System

Fig. 7. Screen shot of General Information display.

Table 1
Data fields in non-clinical section.

Data fields	Examples
Study type	For animals: carcinogenicity study, immunological study, etc.; for human: case reports, clinical trials, etc.
Species	Mouse, rat, dog, human, etc.
Age	Animals: juvenile, adult Humans: children, infants, adults, etc.
Route of administration	Oral, inhalation, dermal, ocular, topical, parenteral, etc.
Dose/concentration	mg, g; mg/kg; mL/kg, etc.
Duration of treatment	Days, weeks, etc.
Frequency of treatment	Per day, per week, etc.
Dosage regimen	Single, repeated, intermittent, etc.
Kinetics data	ADME data, exposure data, etc.
Adverse effects by organ system	Skin, CNS, CVS, Respiratory, Renal, GIT, etc.
Results and conclusion	Specific experimental conditions, recovery findings, etc. (free text)
References	Bibliography

government sources (e.g. JECFA, WHO, EFSA) and other relevant information sources. Hence it is noted "*as available in the public literature*". The information is screened to extract the specific attributes as name of regulatory agency, intake limit, year the limit was assigned and reference with source URL and is organized in tabular format.

3.1.2.2. Non-clinical data. This section of the database compiles documented animal in vivo toxicity studies. The scope of the non-clinical compilation covers pharmacology studies, general toxicity studies, toxicokinetic and non-clinical pharmacokinetic studies, reproduction toxicity studies, carcinogenicity studies and genotoxicity studies. Other non-clinical studies include phototoxicity, immunotoxicity, juvenile animal toxicity and so on. The data were collected from peer-reviewed publications retrieved by systematic and comprehensive literature searches in the electronic bibliographic databases such as PubMed, Embase, Biosis Previews, etc. and factual databases and websites such as International Programme on Chemical Safety (IPCS), the Canadian Centre for Occupational Health and Safety (CCOHS) IPCS Inchem, National Toxicology Program database (NTP), ACToR (Aggregated Computational Toxicology Resource), Hazardous Substances Data Bank (HSDB), European Chemical Substances Information System (ESIS), etc. supplemented by books of toxicological information of chemicals such as Sax's Dangerous Properties of Industrial Materials, Excipients Safety and Toxicity and others.

Table 1 depicts the data fields included in Non-Clinical data section of the pilot version of the STEP database.

3.1.2.3. Clinical data. The clinical data compile safety and tolerability information including epidemiological surveys of populations exposed to a toxic chemical under normal conditions of use, in cases of acute accidental poisoning and in occupational exposure and data from volunteer trials. The main sources of information are international journal articles retrieved from systematic searching of electronic databases, handbooks of toxicity information and poison centers. Human data on the toxicity of chemicals are obviously more relevant to safety evaluation than those obtained from the exposure of experimental animals. However, controlled exposures of man to hazardous or potentially hazardous substances are limited by ethical considerations and information obtained by clinical or epidemiological methods must be relied on. Such data can come from a variety of sources; case reports, case series and toxico surveillance data often contain a high level of detail. However, they are dependent on relatively few data collection systems, for example those of poison control centers. Exposure studies, in particular those involving biomonitoring, provide insight into individual exposure to chemicals in the workplace or environment. However, it is generally difficult to link the often very low levels of chemicals or their metabolites to health outcomes. Conversely, health surveillance may pick up an increased burden of disease but the degree of underlying chemical exposure is often unknown. Volunteer studies have the advantage of well-characterized exposure combined with information on the pharmacokinetic behaviour and/or effects of the compound. Hence wide ranges of sources, from electronic journals to poison centers are explored to collect the human toxicity and safety information.

The data fields for the clinical section are similar to those in the non-clinical section apart from species and age category. The age category for clinical section is mainly grouped into two groups; adults and paediatric. As sometimes it may be more appropriate to collect data over broad age ranges and examine the effect of age as a continuous covariant, the paediatric age group was further divided into different age bands (Preterm new born infants, Term newborn infants (neonates) (0–27 days), Infants and toddlers (1 month–23 months), Pre-school children (2–5 years), School children (6–11 years), Adolescents (12–16 or 18 years)) as per the ICH E11 classification (EMA, 2001). The other attributes are similar to the non-clinical section as displayed in Table 1.

The clinical data section contains data from human studies such as trials, human toxicology studies, poisoning cases, etc. whereas the non-clinical section contains data from all types of toxicological and safety assessment studies. Both the sections are structured in table format to cover the information on experimental design, adverse or beneficial effects and final remarks or conclusion made from the study. As available from individual studies, key attributes are extracted (species, sex and age, study type, dosage form, route of administration, duration and frequency of exposure, dose or concentration, system organ, adverse or beneficial effect and conclusion). The appropriate reference or URL supports each record in the database. Pharmacokinetic studies include information on absorption, distribution, metabolism and excretion of the substance and any active metabolites for both the oral and intended route of administration. The literature may include chapters from reference and textbooks in addition to peer-reviewed journal articles.

3.1.2.4. In vitro data. In vitro toxicity data can be effectively utilized in various aspects of safety/hazard evaluation. Validated in vitro studies can stand alone as independent indicators of risk to human health if a comparable exposure is attained in humans and the in vitro effects correlate with a specific adverse health effect in humans or animals (Frazier, 1992). However there are several issues and challenges, which limit the usefulness of in vitro data (Crump et al., 2010). However, use of in vitro data in risk assessment has great promise towards providing more efficient and cost-effective ways to test chemicals not only to ensure the highest levels of safety, but also to reduce the reliance on animal testing and support innovation. Moreover, it will help to alleviate the large backlog of chemicals that have not been adequately tested.

The in vitro section in the STEP database contains the data from in vitro toxicity testing. The data fields include the detailed information about the protocol used (study type, cell type), the conditions in which the particular experiment was carried out (concentration, controls), the results obtained and the conclusion with free text for comments.

3.1.2.5. Regulatory References. This section of the database provides regulatory information on the excipient such as whether the excipient is subject to any regulation concerning the use and safety of the excipient, or whether the excipient is GRAS listed. It provides links to national information on the regulatory status of the substance, pertinent NRC reports, regulation pertaining to

Table 2
Data fields in the Regulatory References section.

Data fields	Description/example
Excipient name	Benzoic acid
Reference	Bibliography
Source	WHO, EFSA, etc.
URL	Web link to the source
Year	Year of the publication/document (e.g. 2012)

the excipients as published in Federal Register, worldwide food additive status (e.g. JECFA, WHO), assessment reports prepared by regulatory agencies, scientific opinions and advice from regulatory and government agencies (e.g. EFSA, European commission, FDA). It provides the information on accepted uses in foods and licensed pharmaceuticals where known. However the status of excipients varies from one nation to another and appropriate regulatory bodies need to be consulted for guidance. With this database, end-users will have access to a single source of information, making it easier to stay up to date and comply with regulations in an efficient and prompt manner. The information is extracted and organized as per the data fields shown in Table 2.

3.1.2.6. Reviews section. The review articles discussing one or more facets of the toxicology/safety of the excipient are compiled in this section of the database. Most of these articles do not contain the specific information that can be extracted as per the data fields into the database. However, they do provide useful information about the safety or toxicity of the excipient or group of related excipients. Some articles may contain a complete, detailed description of the toxicity of an excipient, others may address only a particular aspect of the toxicity and others may only list the excipient in a general discussion of the toxicity of a class of compounds. The list of such references is made available with the URL link to the source.

3.1.2.7. Export data section. Generating a report of the retrieved data was one of the essential requirements for database. Export interface is hence provided to allow the users to create a flexible, need based information reports that can be integrated it into different applications. Fig. 8 represents an example of the generated report. The data fields are provided on the export interface for the users to be able to filter/select the fields as per their needs, generate report and save it in required format for further analysis. For instance, to generate the report for dose and results information, the user can select the "dose and result" check boxes and the report will be created in required format for further statistical analysis.

4. Data collection and curation

Data gathering and curation processes are carried out in parallel with the development of the database application and will continue throughout the duration of the project. In line with the priorities set in the database development, data collection commenced with 10 prioritized excipients so that the content of the pilot database is sufficient to demonstrate an ability to support searching based on queries about excipients and combination of excipients as well as different types of adverse effects and safety data, and combinations of these parameters. Also the large number of excipients and the limited toxicity information for many of these have already been identified as a driver for needing to set priorities for additional testing (Judson et al., 2009). Because the list of excipients is so large, it was necessary to prioritize which excipients will be included into the STEP database. In this pilot

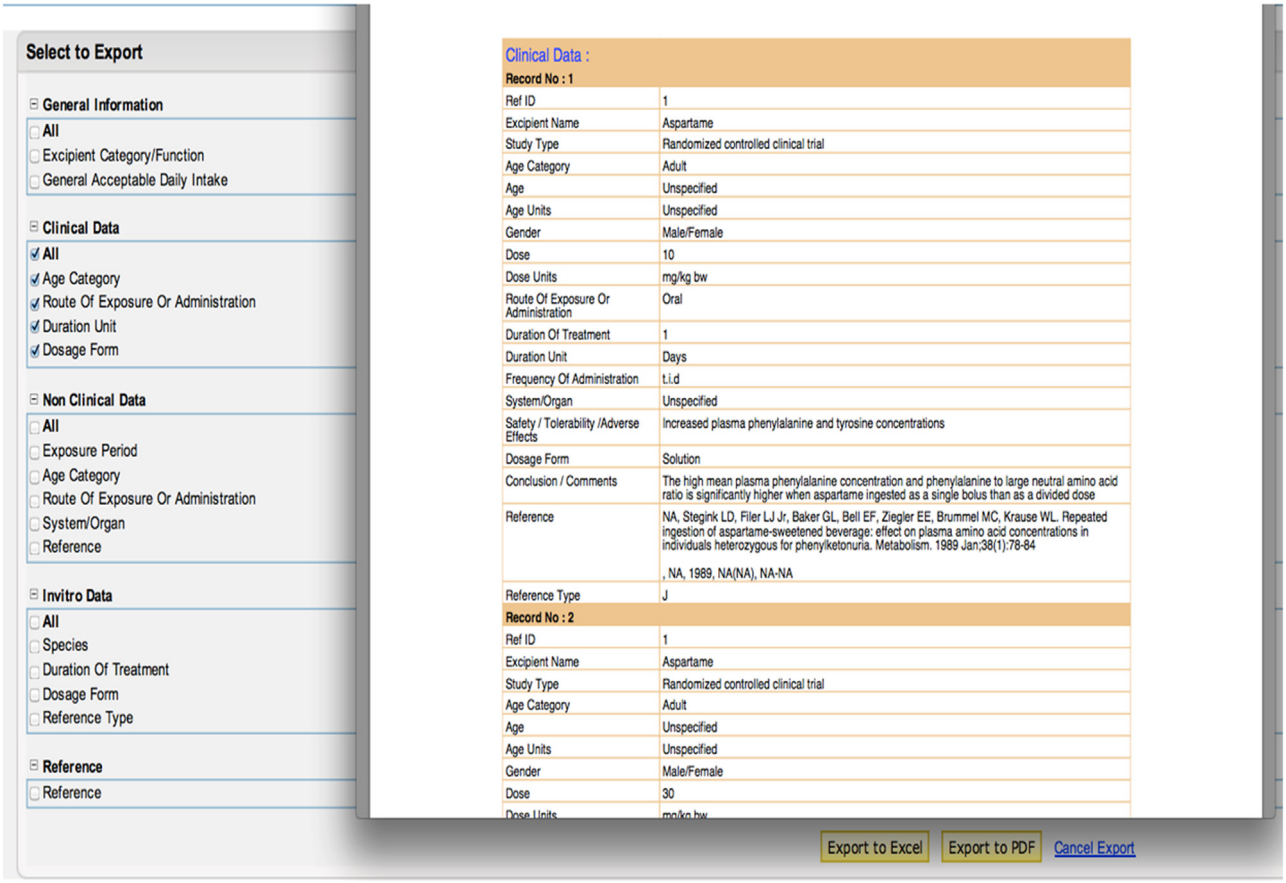


Fig. 8. Screen Shots of Export Page & Report generated from the STEP database.

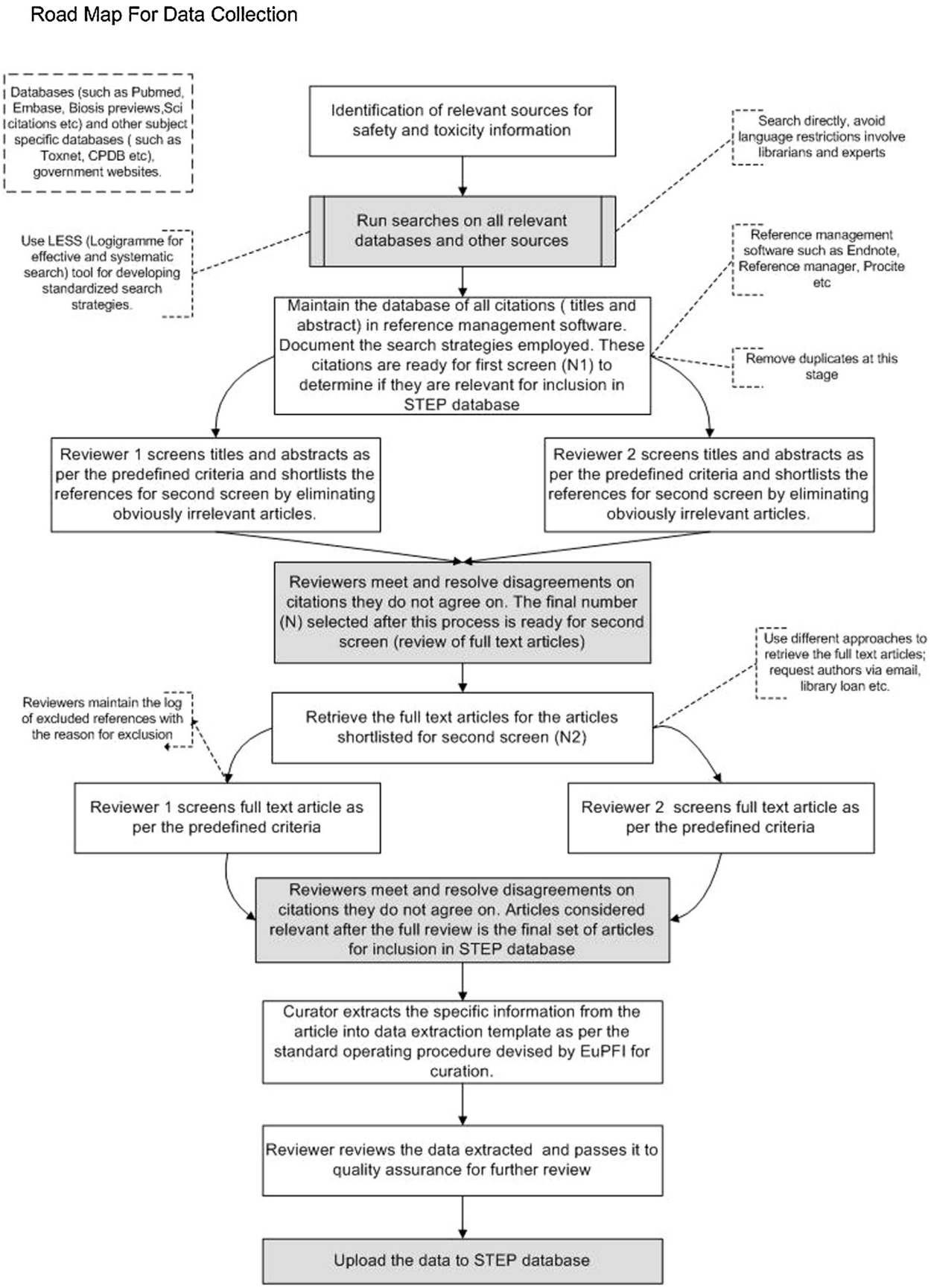


Fig. 9. Road map for data collection.

phase of database development, both excipients, for which significant toxicology information is limited and well-characterized excipients are included. The well-characterized excipients were important because they allow for the development of predictive models for prioritization of the remaining largely uncharacterized excipients. The following selection criteria for excipients of interest were considered:

1. Excipients known to be toxic/have safety issues in general
2. Frequency of occurrence as contaminants or toxic in paediatrics (where appropriate)
3. Evidence in literature of toxicity in paediatrics.

The above criteria were applied to identify the excipients for inclusion in the STEP database. The excipients were short-listed/prioritized through polling within the Eu & US PFI members as listed below.

List of prioritized excipients for pilot database development

1. Propylene glycol (PG)
2. Ethanol
3. Polysorbate 80
4. Benzyl alcohol
5. Parabens (propyl, methyl, ethyl and butyl)
6. Benzalkonium chloride
7. Aspartame
8. Sorbitol
9. Benzoic acid
10. Sodium benzoate

The aim was to use a selection of prioritized excipients for developing data gathering methodology and for further validation of the methodology. The peer-reviewed, published literature was the starting point for the data collection. After extensive research on data collection procedures, a methodology was devised to identify the relevant information sources (Medline, Biosis Previews, International Pharmaceutical Abstracts, Chemical Abstract, CINAHL, etc.) and to standardize resource specific search strategies to retrieve relevant references from these sources (Brandys and Salunke, 2012; Vaconsin et al., 2012).

Fig. 9 depicts the complete roadmap for data collection. The defined search strategies were run on selected sources. There was no time limit on the search query. All references were compiled in the reference management software. Duplicate references were removed at this stage and the unique references were thoroughly screened for relevancy as per a predefined inclusion and exclusion criteria (Salunke and Tuleu, 2010). Full texts were obtained where needed and read to determine if they are relevant for inclusion in the database. The studies excluded were those that used a combination of excipients where the effect of the single excipient could not be defined and/or those whose references were not traceable. The data from unpublished studies were included only if the source was well identified.

In order to populate the STEP database, a data extraction process was required. Manual extraction has been proven to be both time consuming and labour/cost intensive and hence was decided to outsource this to external companies. Several curation service providers offering manual data extraction with sufficient capacity, good quality standards and assurance of confidentiality were evaluated for use in this project. A Data Curation manual was prepared to clearly define and describe the data extraction procedure, breaking out each step in detail so that the curator understands the content that needs curation. The manual included protocols, policies and procedures; the data collection instrument; and a listing of all the data elements and their full definitions. To further optimize/expedite the data curation process, a standardized template containing some predefined information of the experiment; and an automatic check for formal errors was developed. This allowed reducing the possible typing errors as well as facilitating and speeding up the data entry. Also proper training on the protocol and procedures, data sources, data collection instrument, and most importantly data definitions was provided to curators. Thesauri/ontologies were developed for excipient names, effects, species, exposure routes, and doses to maintain the consistency in the data entered and prevention of interpretive errors. The key attributes for extraction include demographic information (species, age and gender), experimental design/protocol information (dosage form, study type, dose, route of administration, duration of exposure, frequency of administration) and results (adverse or beneficial effects, specific remarks or conclusion of the study). The aim was to collect a full set of information for all prioritized excipients but this was not always possible. The reasons differed from each other, but in most cases some information were not reported in the literature. Papers in a foreign language were translated to English where possible. If no translation of the full paper was available, the English language abstract, was used to extract the information. In such case, this was notified in the "Comments" field in the database. References with incomplete data were also included in the database to give the widest possible picture of the studies. In such cases, information was noted as "unspecified".

While it is considered worthwhile to build a database compiling safety and toxicity information already in the public domain, it was recognized that the data in the collected preclinical toxicity reports of pharmaceutical companies represents a potential data resource. Efforts are underway to gather more data from legacy reports, unpublished documents and in-house non-confidential data from pharmaceutical companies. In the long term one of the aims of the STEP database would be to promote the sharing of data from individual corporate entities to avoid future duplication of studies, lead to more efficient drug development processes and at the same time contribute to the reduction of animal studies. The pilot database was hence populated with some data donated from industry, to demonstrate a mechanism for sharing such data, and the potential benefits of doing so.

The quality of the data entry into the database is an essential element and needs a thorough quality assurance to ensure there are no errors in the database and that the required levels of precision and recall are achieved or exceeded. According to this, strict quality control procedures were implemented at the collection level, involving proof-reading of all output documents and comprehensive checking of selected curated information against the original sources.

4.1. Primary

Primary quality assurance is performed when a curator initially enters details collected from the source information. A Curator is an experienced scientist with Mpharm/PhD degree and knowledge of pharmacology and toxicology. At this time the individual curator reviewing the study, examined all data reported, to determine if the criteria required for the study to be included in the database were met and the information was then extracted as per the template. A senior curator secondarily reviewed the extracted information.

4.2. Secondary

Secondary quality assurance for these data is considered to be the review and comparison of the data extracted, and entered into an excel spreadsheet, with the data presented in the actual reference. The secondary review process was conducted by reviewers, but not by the person who initially entered the data. This quality assurance effort was conducted on every record entered into the database. In addition, reviewers assist the database team in qual-

Table 3
The STEP database technical specifications.

Project server	Connecting	Server software			
		Operating system	Web application server	Database server	Application language
	JDBC	Red Hat Enterprise Linux	Tomcat Apache Server	Postgre SQLOracle	Dot net

ity assurance by noting any entry data that is found to have errors during their day-to-day activities. The references are reallocated for review to curators to achieve an error <0.1%.

4.3. Tertiary

Tertiary quality assurance is implemented following a completion of secondary review. The methods used in this third review include random visual check of data entries, several methods of sorting by data fields to detect inconsistencies, and further investigation into records that contain these inconsistencies. The IT personnel then uploaded these data into the database. Once the review of all the available fields is completed, through sufficient quality checks, the Excel spreadsheet is ported into Oracle.

The overall system of data curation and quality control was managed manually currently as an interim solution and in future it is planned to automate at least some of the functions.

5. Software development

The database application has been developed in Oracle, a Relational Database Management System (RDBMS), and the web front-end uses Java (Table 3). It is designed in a way to allow future extension.

All database connections and application logic are completely transparent to the Web browser. Access to the database is secured using email and the SSL protocol. This will ensure that all data updates are secure and will enable appropriate access control for individual users. At the registration of user access to the database EuPFI will allow the use of database. Any kind of requests made by the end-users via Web Browsers hit the Web Server first. The request is then processed by Application Interface deployed in the server. Application Interface uses database connection and fetches the required information and provides multiple ways of fetching and storing the data in the database including Alerts, Query Search, Export, user Administration, etc. The interface available is fully web-based. The presentation layer is based on html/css documents. The Javascript language extended to the popular query language is responsible for the interface operations (Fig. 10)

6. Discussion

Over the past decade a dramatic increase has occurred in the recognition and concern for children as a potential susceptible population for exposure to inappropriate excipients. Developing medicines for children is more complex than developing medicines for adults and therefore it is important that a complete risk assessment approach to using excipients is considered. Evidence based risk assessments would enable selection of more appropriate excipients. However, evidence based toxicity assessment of excipients requires a body of evidence to be retrospectively assessed: data about the adverse effects associated with excipients and the expo-

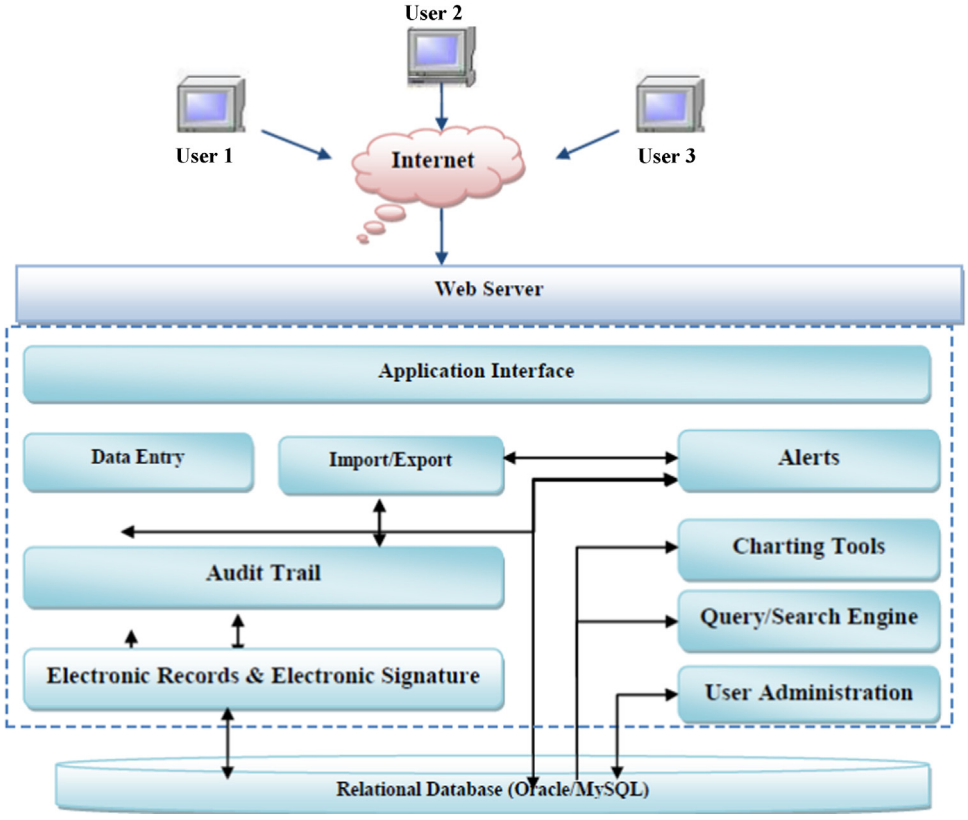


Fig. 10. The STEP database software operation.

sure levels associated with those effects. As the body of knowledge on children's health and risk factors expands rapidly, there is an increasing need for systematic compilation of knowledge for further application (screening/prioritization of the excipients to be used in paediatric formulation, risk assessment process, etc.). Currently there is no comprehensive resource that provides age specific data. Review of excipient-specific and age group-specific data may reveal a specific age group of a particular concern from a toxicokinetic perspective. These resources are crucial for the toxicologists and regulators involved in the risk assessment of chemicals, which necessitates access to all the relevant literature, and the capability to search across toxicity databases using both biological and chemical criteria. Different possibilities to address this problem are currently being analysed. Existing public resources are difficult to search due to free text format; hence disaggregating the information to make it searchable in the STEP database would be an opportunity to add value to existing knowledge.

A pilot web-based tool is developed for storage and exchange of safety and toxicological data for application in paediatric drug development. It is a unique database that combines clinical and non-clinical data of excipients and allows for easy searching and filtering of the data as per needs. It is not however, meant to serve directly as a data analysis/computational tool; it does not include any software that would allow statistical calculations. The STEP database does not critically review or evaluate studies for scientific quality. For example, the STEP database generally does not evaluate study design strengths, weaknesses and limitations, whether the evidence it reports supports conclusions in a study, or whether there is conflicting evidence from other studies. The STEP database does not derive quantitative estimates for risk assessment purposes and does not integrate findings among studies. Listing of a study in the STEP database does not represent an endorsement as to the scientific quality of the study. The database is provided as a free Internet resource, thus ensuring easy and cost effective access for all end-users (academic, industrial and regulatory toxicologists, pharmacologists, clinicians, researchers, regulators and others). It is an important contribution to the rational choices of excipients for paediatric and neonatal drug formulations. The development of the STEP database will facilitate studies in this area and will help identify those excipients that take priority for evaluation in newborns and children and would form the basis of where the research efforts should be directed.

Throughout the process of the database development, the end users were consulted and so the user requirements and testing has been integral to the design. The schema for the database has been developed in an iterative manner. The first draft was created by EuPFI and was implemented into the web based STEP database by GVK BIO. Several improvements were made during the course of a pilot study based on the feedback from participating EuPFI members and domain experts.

To date, the STEP database contains a full set of data regarding the selected 10 excipients, covering the general information link to physical chemical properties, and detailed toxicity and safety information. About 700 references of clinical data, 1200 data records of nonclinical data, 500 records from in vitro experiments, 288 references related to regulatory information and other interesting links and finally 233 toxicological and clinical reviews were curated in this first version of the database. These data can be easily retrieved and exported in different formats to facilitate data management and analyses.

The pilot data extraction study highlighted the need for a consensus to be reached on how the database schema should be populated. The questions and answers generated during the pilot study have been used to define the data entry guidelines to ensure consistency among the curators entering the data. The EuPFI members carried out the quality assessments and consistency checks on the extracted data.

Key issues that the pilot database development have highlighted and need further efforts are extraction of quantitative data into tabular, computable formats and mapping of data from multiple sources onto standard terms and using predefined vocabularies or ontologies. The design should involve compromises between completeness and comprehensiveness and curated versus unfiltered content and has a query language that exploits this type of information. The largest challenge is the desire of the users to have more detailed information on many excipients. With these issues have come concerns about how to provide the long-term maintenance and operations of this infrastructure. To enable the long-term sustainability of the STEP database, it is imperative to develop a strategic approach that involves users. There is a genuine need to involve the users, since they are served by this infrastructure and any decisions about alternative funding models might affect their work and research. The STEP database needs periodic solicitation of feedback to ensure the infrastructure and services continue to serve the needs of the scientific community. The success of the database will depend heavily on continuous input from the industry, academics, hospitals, regulatory agencies and related communities about their needs and priorities. Hence, the feedback on the utility of the database and how it can be improved in subsequent releases is encouraged. More effort and coordination is needed in order to fully address the issue of the STEP database sustainability and ensure this valuable resource is available in the future to continue supporting the excipients research and fostering innovation.

In conclusion, to the best of our knowledge, the presented database is the only publicly available readily accessible source presenting the safety and toxicity information of excipients with paediatric element included. It is anticipated that this source will prove to be a useful platform for data retrieval and management and data exchange of excipient safety information. It will serve as an excellent source of diligently curated data for development of safer medicines for children. For instance, it has already shown to be of value during the development of "1st excipient monograph of propylene glycol" by ESNEE (Vaconsin et al., 2013). The intention is for the STEP database to be useful to those people responsible for developing medicines designed for children, as well as for those responsible for manipulating adult licensed medicines for off-label use in children. However, in addition to the paediatric data, the database includes adult, animal, and in vitro data and so also can be for other purposes (e.g. selection of species for toxicity studies, toxicity prediction studies, etc.). Therefore, the users are encouraged to consider this as a general database that could be applicable for other areas of toxicology. Such a database would serve as a valuable tool for archiving and also facilitate exchange of data from any safety and toxicological studies such as ESNEE project (Turner and Storme, 2012). The STEP database needs periodic solicitation of feedback to ensure the infrastructure and services continue to serve the needs of the scientific community. The success of the database will depend heavily on continuous input from the industry, academics, hospitals, regulatory agencies and related communities about their needs and priorities. The feedback on the utility of the database and how it can be improved in subsequent releases is encouraged.

Following the pilot described in this paper a follow-on long term project has started. The focus of the ongoing project is to increase the number of excipients in the existing database so that a database large enough to be of practical research use will be available. Following to that, database maintenance, updating and expansion will continue. The database is available to a large circle of the users (toxicologists, clinicians, regulators, researchers, etc.) and feedback will be collected on a regular basis to capture the needed additional changes or requirements. Furthermore, currently the only source of information is published literature, however it is well

acknowledged that the vast amount of information is held within the drug companies. Hence companies are encouraged to step forward towards a collaborative effort to contribute their in-house non-confidential data into the STEP database to avoid the duplication of the efforts and save money and time involved in toxicity testing.

This paper is the second in a series of articles presenting the development of the pilot STEP database. Subsequent articles will present the challenges associated with the development and methodology developed to overcome those, along with the beta testing feedback from the end-users.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007–2013 under grant agreement no. 261060.

The authors would like to thank GVK BIO database development team for their contribution to the end-to-end development of database including designing the database from the early stages of developing database and extensive modification due to changes in requirement. We would like to acknowledge the support of the Eu and US PFI members and beta version testers for their contribution to the development and testing of the database. The authors wish to thank Patricia Fowler in the proof-reading of the manuscript and Pertti (Bert) J. Hakkinen, Ph.D. Acting Head, Office of Clinical Toxicology Specialized Information Services National Library of Medicine, for his advice and technical support.

References

Allegaert, K., Vanhaesebrouck, S., Kulo, A., Cosaert, K., Verbesselt, R., Debeer, A., de Hoon, J., 2010. Prospective assessment of short-term propylene glycol tolerance in neonates. Arch. Dis. Child. 95, 1054–1058.

Balistreri, W.F., Farrell, M.K., Bove, K.E., 1986. Lessons from the E-Ferol tragedy. Pediatrics 78, 503–506.

Brandys, B., Salunke, S., 2012. The STEP (Safety and Toxicity of Excipients for Paediatrics) Database. In: Special Libraries Association (SLA) Annual Conference, Chicago, IL, Available at http://dbiosla.org/events/conf_past/Chicago/contr_papers2012.html (accessed 08.08.13).

Carter, 2011. Risk assessment for excipients for enhanced patient safety. Pharm. Technol. 35, 29–33.

CHMP, 2006. Reflection paper: Formulations of choice for the paediatric population. EMEA/CHMP/PEG/194810/2005. Available at.

Crump, K.S., Chen, C., Louis, T.A., 2010. The future use of in vitro data in risk assessment to set human exposure standards: challenging problems and familiar solutions. Environ. Health Perspect. 118, 1350–1354.

EMA, 2001. ICH Topic E11 Guideline on Clinical Investigation of Medicinal Products in the Paediatric Population CPMP/ICH/2711/99.

Ernest, T.B., Elder, D.P., Martini, L.G., Roberts, M., Ford, J.L., 2007. Developing paediatric medicines: identifying the needs and recognizing the challenges. J. Pharm. Pharmacol. 59, 1043–1055.

Fabiano, V., Mameli, C., Zuccotti, G.V., 2011. Paediatric pharmacology: remember the excipients. Pharmacol. Res. 63, 362–365.

FDA, 2011. Drug Safety Communication: Serious health problems seen in premature babies given Kaletra (lopinavir/ritonavir) oral solution, Available at http://www.fda.gov/Drugs/DrugSafety/ucm246002.htm (accessed 17.07.13).

Frazier, J.M., 1992. In-Vitro Toxicity Testing: Application to Safety Evaluation, First edition. Informa healthcare, New York.

Guengerich, F., MacDonald, J., 2007. Applying mechanisms of chemical toxicity to predict drug safety. Chem. Res. Toxicol. 20, 344–369.

Judson, R., Richard, A., Dix, D.J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., Tan, S., Carpenter, T., Smith, E., 2009. The toxicity data landscape for environmental chemicals. Environ. Health Perspect. 117, 685–695.

Kulo, A., de Hoon, J.N., Allegaert, K., 2012. The propylene glycol research project to illustrate the feasibility and difficulties to study toxicokinetics in neonates. Int. J. Pharm. 435, 112–114.

Lass, J., Naelapää, K., Shah, U., Käär, R., Varendi, H., Turner, M.A., Lutsar, I., 2012. Hospitalised neonates in Estonia commonly receive potentially harmful excipients. BMC Pediatr. 12, 136.

Maldonado, S., Schaufelberger, D., 2011. Pediatric formulations. Am. Pharm. Rev., Available at http://www.americanpharmaceuticalreview.com/Featured-Articles/37186-Pediatric-Formulations/ (accessed 17.07.13).

Nunn, T., Williams, J., 2005. Formulation of medicines for children. Br. J. Clin. Pharmacol. 59, 674–676.

Rowe, R.C., Sheskey, P.J., Cook, W.G., Fenton, M.E., 2012. Handbook of Pharmaceutical Excipients, Seventh edition. Pharmaceutical Press.

Shehab, N., Lewis, C.L., Streetman, D.D., Donn, S.M., 2009. Exposure to the pharmaceutical excipients benzyl alcohol and propylene glycol among critically ill neonates. Pediatr. Crit. Care Med. 10, 256–259.

Salunke, S., Tuleu, C., 2010. Pilot database addressing safety and toxicity aspects of excipients for paediatric medicines development. In: 2nd European Paediatric Formulation Initiative (EuPFI) Conference, Available at http://www.eupfi.org/downloads/2ndconferenceAbstractsforPublication.pdf (accessed 17.07.13).

Salunke, S., Giacoia, G., Tuleu, C., 2012. The STEP (Safety and Toxicity of Excipients for Paediatrics) database. Part 1—A need assessment study. Int. J. Pharm. 435, 101–111.

Vaconsin, P., Storme, T., Salunke, S., Turner, M., Brion, F., Rieutord, A., 2012. ESNEE project: strategy for the extensive review of the literature on excipient toxicities. Eur. J. Hosp. Pharm.: Sci. Pract. 19, 183.

Vaconsin, P., Storme, T., Salunke, S., Masset, D., Nunn, A., Metsvaht, T., Carleer, J., Turner, M., Rieutord, A., 2013. GRP-001 1St ESNEE Excipient Monograph: information needed to formulate, prepare and prescribe medicines for neonates containing propylene glycol as an excipient. Eur. J. Hosp. Pharm.: Sci. Pract. 20, A1, Poster available at http://www.eahp.eu/sites/default/files/grp-001.pdf (accessed 17.07.13).

Tomasulo, 2002. ChemIDplus – super source for chemical and drug information. Med. Ref. Serv. Q. 21, 53–59.

Turner, M.A., Storme, T., 2012. European Study for Neonatal Excipient Exposure (ESNEE). Eur. J. Hosp. Pharm.: Sci. Pract. 19, 67.

Ursino, M.G., Poluzzi, E., Caramella, C., De Ponti, F., 2011. Excipients in medicinal products used in gastroenterology as a possible cause of side effects. Regul. Toxicol. Pharmacol. 60, 93–105.

WHO, 1987. Principles for the safety assessment of food additives and contaminants in food. In: Environmental Health Criteria no. 70. World Health Organisation, Geneva.